On the Higher-Order Derivatives of Spectral Functions: Two Special Cases

Hristo S. Sendov*

February 1, 2008

Abstract

In this work we use the tensorial language developed in [8] and [9] to differentiate functions of eigenvalues of symmetric matrices. We describe the formulae for the k-th derivative of such functions in two cases. The first case concerns the derivatives of the composition of an arbitrary differentiable function with the eigenvalues at a matrix with distinct eigenvalues. The second development describes the derivatives of the composition of a separable symmetric function with the eigenvalues at an arbitrary symmetric matrix. In the concluding section we re-derive the formula for the Hessian of a general spectral function at an arbitrary point. Our approach leads to a shorter, streamlined derivation than the original in [6]. The language we use, based on the generalized Hadamard product, allows us to view the differentiation of spectral functions as a routine calculus-type procedure.

1 Introduction

We say that a real-valued function F, on a symmetric matrix argument, is spectral if it has the following invariance property:

$$F(UXU^T) = F(X),$$

for every symmetric matrix X in its domain and every orthogonal matrix U. The restriction of F to the subspace of diagonal matrices defines (almost) a function $f(x) := F(\text{Diag } x)$ on a vector argument $x \in \mathbb{R}^n$. It is easy to see that $f : \mathbb{R}^n \to \mathbb{R}$ has the property

$$f(x) = f(Px) \text{ for any permutation matrix } P \text{ and any } x \in \text{domain } f.$$

---

*Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario, Canada N1G 2W1. Email: `hssendov@uoguelph.ca`. Research supported by NSERC.

We call such functions *symmetric*. It is not difficult to see that $F(X) = f(\lambda(X))$, where $\lambda(X)$ is the vector of eigenvalues of X. An important subclass of spectral functions is obtained when $f(x) = g(x_1) + \cdots + g(x_n)$ for some function g on one real variable. We call such symmetric functions *separable* and their corresponding spectral functions will be called *separable spectral* functions.

In [4] an explicit formulae for the gradient the spectral function F in terms of the derivatives of the symmetric function f was given:

$$(1) \qquad \nabla(f \circ \lambda)(X) = V\big(\mathrm{Diag}\,\nabla f(\lambda(X))\big)V^T,$$

where V is any orthogonal matrix such that $X = V\big(\mathrm{Diag}\,\lambda(X)\big)V^T$ is the *ordered* spectral decomposition of X. In [6] a formula for the Hessian of F was given, whose structure appeared quite different than the one for the gradient. In this work we generalize the work in [4] and [6] by proving the following formula for the k-th derivative of a spectral function

$$(2) \qquad \nabla^k(f \circ \lambda)(X) = V\Big(\sum_{\sigma \in P^k} \mathrm{Diag}^\sigma \mathcal{A}_\sigma(\lambda(X))\Big)V^T,$$

where again $X = V(\mathrm{Diag}\,\lambda(X))V^T$. The sum is taken over all permutations on k elements. (The role of the permutations is just as a convenient tool for enumerating the maps $\mathcal{A}_\sigma(x)$.) The precise meaning of the operators Diag^σ, generalizing the Diag operator, is explained in the next section, see Formula (4). The main thing to keep in mind about the formula is that the maps $\mathcal{A}_\sigma(x)$ depend only on the partial derivatives of $f(x)$, up to order k, and do not depend on the eigenvalues. In this sense the process of differentiating $f \circ \lambda$ leaves the eigenvalues unscathed, since the only way in which they participate in the formula above is through the compositions $\mathcal{A}_\sigma(\lambda(X))$ and the conjugation by the orthogonal matrix V.

We show that Formula (2) holds in two general subcases. It holds when f is a k-times differentiable function, not necessarily symmetric, and X is a matrix with distinct eigenvalues. It also holds when f is a separable symmetric function and X is an arbitrary symmetric matrix. We give an easy recipe for computing the maps $\mathcal{A}_\sigma(x)$ in the above two cases.

In addition, we show that in the case when f is a k-times *continuously* differentiable, separable, symmetric function, Formula (2) can be significantly simplified. In that case, all the maps $\mathcal{A}_\sigma(x)$ coincide, that is $\mathcal{A}_{\sigma_1}(x) = \mathcal{A}_{\sigma_2}(x)$ for any two permutations σ_1, σ_2 on k elements.

Finally, in the last section, we re-derive the formula for the Hessian of a general spectral function at an arbitrary symmetric matrix. Our approach leads to a shorter, more streamlined derivation than the original derivation in [6].

The language that we use, based on the generalized Hadamard product, allows us to differentiate Formula (2) just like one would expect: writing the differential quotient and taking the limit as the perturbation goes to zero. This gives a clear view of where the different pieces in the differential come from and give the process a routine Calculus-like flavour.

In the next section, we give all the necessary notation, definitions, and background results to make the reading of this work self-contained.

2 Notation and background results

We use pretty much the same notation as in the preceding two papers [8] and [9]. We will briefly summarize it here for completeness and will try to make the reading of this part independent.

By S^n, O^n, and P^n we will denote the set of all $n \times n$ real symmetric, orthogonal, permutation matrices respectively. By M^n will be denoted the real Euclidean space of all $n \times n$ matrices with inner product $\langle X, Y \rangle = \operatorname{tr}(XY^T)$. For $A \in S^n$, $\lambda(A) = (\lambda_1(A), ..., \lambda_n(A))$ will be the vector of its eigenvalues ordered in nonincreasing order. By $\mathbb{N}_k$ we will denote the set $\{1, 2, ..., k\}$. For any vector x in $\mathbb{R}^n$, $\operatorname{Diag} x$ will denote the diagonal matrix with the vector x on the main diagonal, and $\operatorname{diag}: M^n \to \mathbb{R}^n$ will denote its conjugate operator, defined by $\operatorname{diag}(X) = (x_{11}, ..., x_{nn})$. By $\mathbb{R}^n_\downarrow$ we denote the cone of all vectors x in $\mathbb{R}^n$ such that $x_1 \geq x_2 \geq \cdots \geq x_n$. Denote the standard basis in $\mathbb{R}^n$ by $e^1, e^2, ..., e^n$. For a permutation matrix $P \in P^n$ we say that $\sigma : \mathbb{N}_n \to \mathbb{N}_n$ is its corresponding permutation map and write $P \leftrightarrow \sigma$ if for any $h \in \mathbb{R}^n$ we have $Ph = (h_{\sigma(1)}, ..., h_{\sigma(n)})^T$ or, in other words, $P^T e^i = e^{\sigma(i)}$ for all $i = 1, ..., n$. The symbol δ_{ij} will denote the Kroneker delta. It is equal to one if $i = j$ and zero otherwise.

Any vector $\mu \in \mathbb{R}^n$ defines a partition of $\mathbb{N}_n$ into disjoint *blocks*, where integers i and j are in the same block if, and only if, $\mu_i = \mu_j$. In general, the blocks that μ determines need not contain consecutive integers. We agree that the block containing the integer 1 will be the first block, I_1, the block containing the smallest integer that is not in I_1 will be the second block, I_2, and so on. The number r will denote the number of blocks in the partition. Let ι_l denote the largest integer in I_l for all $l = 1, ..., r$. For any two integers, $i, j \in \mathbb{N}_n$ we will say that they are *equivalent (with respect to μ)* and write $i \sim j$ (or $i \sim_\mu j$) if $\mu_i = \mu_j$, that is, if they are in the same block. Two k-indexes $(i_1, ..., i_k)$ and $(j_1, ..., j_k)$ are called *equivalent* if $i_l \sim j_l$ for all $l = 1, 2, ..., n$, and we will write

$$(i_1, ..., i_k) \sim (j_1, ..., j_k).$$

A *k-tensor, T, on $\mathbb{R}^n$* is a map from $\mathbb{R}^n \times \cdots \times \mathbb{R}^n$ (k-times) to $\mathbb{R}$ that is linear in each argument separately. Denote the set of all k-tensors on $\mathbb{R}^n$ by $T^{k,n}$. The value of the k-tensor at $(h_1, ..., h_k)$ will be denoted by $T[h_1, ..., h_k]$. The tensor is called *symmetric* if for any permutation, σ, on $\mathbb{N}_k$ it satisfies $T[h_{\sigma(1)}, ..., h_{\sigma(k)}] = T[h_1, ..., h_k]$, for any $h_1, ..., h_k \in \mathbb{R}^n$. Given a vector $\mu \in \mathbb{R}^n$, a tensor $T \in T^{k,n}$ is *μ-symmetric* if for any permutation $P \in P^n$, such that $P\mu = \mu$, we have $T[Ph_1, ..., Ph_k] = T[h_1, ..., h_k]$, for any $h_1, ..., h_k \in \mathbb{R}^n$. A k-tensor valued map, $\mu \in \mathbb{R}^n \to \mathcal{F}(\mu) \in T^{k,n}$, is *$\mu$-symmetric* if for every $\mu \in \mathbb{R}^n$ and permutation matrix P we have $\mathcal{F}(P\mu)[Ph_1, ..., Ph_k] = \mathcal{F}(\mu)[h_1, ..., h_k]$, for any $h_1, ..., h_k \in \mathbb{R}^n$. The tensor is called is called *μ-block-constant* if $T^{i_1 \cdots i_k} = T^{j_1 \cdots j_k}$ whenever $(i_1, ..., i_k) \sim (j_1, ..., j_k)$. A k-tensor valued map, $\mu \in \mathbb{R}^n \to \mathcal{F}(\mu) \in T^{k,n}$, is *block-constant* if $\mathcal{F}(\mu)$ is μ-block-constant for every μ. Clearly, every μ-block-constant tensor is μ-symmetric. By $T[h]$ we denote the $(k-1)$-tensor on $\mathbb{R}^n$ given by $T[\cdot, ..., \cdot, h]$. Similarly for $T[M]$, if T is a k-tensor on M^n and $M \in M^n$. The following easy lemma was proved in [9].

Lemma 2.1 *If a k-tensor valued map, $\mu \in \mathbb{R}^n \to T(\mu) \in T^{k,n}$, is μ-symmetric and differentiable, then its differential, $\nabla T(\mu)$, is also μ-symmetric.*

For each permutation σ on $\mathbb{N}_k$, we define *σ-Hadamard* product between k matrices to be a

k-tensor on $\mathbb{R}^n$ as follows. Given any k basic matrices $H_{p_1 q_1}, H_{p_2 q_2},...,H_{p_k q_k}$

$$(H_{p_1 q_1} \circ_\sigma H_{p_2 q_2} \circ_\sigma \cdots \circ_\sigma H_{p_k q_k})^{i_1 i_2 ... i_k} = \begin{cases} 1, & \text{if } i_s = p_s = q_{\sigma(s)}, \forall s = 1, ..., k, \\ 0, & \text{otherwise.} \end{cases}$$

Extend this product to a multi-linear map on k matrix arguments:

$$(3) \qquad (H_1 \circ_\sigma H_2 \circ_\sigma \cdots \circ_\sigma H_k)^{i_1 i_2 ... i_k} = H_1^{i_1 i_{\sigma^{-1}(1)}} \cdots H_k^{i_k i_{\sigma^{-1}(k)}}.$$

Notice that when $k = 1$ we have $\circ_{(1)} H = \text{diag} \, H$. Let T be an arbitrary k-tensor on $\mathbb{R}^n$ and let σ be a permutation on $\mathbb{N}_k$. We define $\text{Diag}^\sigma T$ to be a $2k$-tensor on $\mathbb{R}^n$ in the following way

$$(4) \qquad (\text{Diag}^\sigma T)^{i_1 ... i_k}_{j_1 ... j_k} = \begin{cases} T^{i_1 ... i_k}, & \text{if } i_s = j_{\sigma(s)}, \forall s = 1, ..., k, \\ 0, & \text{otherwise.} \end{cases}$$

When $k = 1$ we have $\text{Diag}^{(1)} x = \text{Diag} \, x$ for any $x \in \mathbb{R}^n$. Any $2k$-tensor, T, on $\mathbb{R}^n$ can naturally be viewed as a k-tensor on M^n in the following way

$$T[H_1, ..., H_k] = \sum_{p_1, q_1 = 1}^{n} \cdots \sum_{p_k, q_k = 1}^{n} T^{p_1 ... p_k}_{q_1 ... q_k} H_1^{p_1 q_1} \cdots H_k^{p_k q_k}.$$

Define dot product between two tensors in $T^{k,n}$ in the usual way:

$$\langle T_1, T_2 \rangle = \sum_{p_1, ..., p_k = 1}^{n} T_1^{p_1 ... p_k} T_2^{p_1 ... p_k}.$$

We define an action (called *conjugation*) of the orthogonal group O^n on the space of all k-tensors on $\mathbb{R}^n$. For any k-tensor, T, and $U \in O^n$ this action will be denoted by $UTU^T \in T^{k,n}$:

$$(5) \qquad (UTU^T)^{i_1 ... i_k} = \sum_{p_1 = 1}^{n} \cdots \sum_{p_k = 1}^{n} \left(T^{p_1 ... p_k} U^{i_1 p_1} \cdots U^{i_k p_k} \right).$$

We showed in [8] that this action is norm preserving and associative: $V(UTU^T)V^T = (VU)T(VU)^T$ for all $U, V \in O^n$.

The Diag^σ operator, the σ-Hadamard product, and conjugation by an orthogonal matrix are connected by the following formula, see [8].

Theorem 2.2 *For any k-tensor T, any matrices $H_1,...,H_k$, any orthogonal matrix V, and any permutation σ in P^k we have the identity*

$$(6) \qquad \langle T, \tilde{H}_1 \circ_\sigma \cdots \circ_\sigma \tilde{H}_k \rangle = \left(V(\text{Diag}^\sigma T)V^T \right)[H_1, ..., H_k],$$

where $\tilde{H}_i = V^T H_i V$, $i = 1, ..., k$.

We will also need the following lemma from [8].

Lemma 2.3 *Let T be any $2k$-tensor on R^n, $U \in O^n$, and let H be any matrix. Then, the following identity holds.*

$$U(T[U^T H U])U^T = (UTU^T)[H].$$

Given a permutation σ on $\mathbb{N}_k$ we can naturally view it as a permutation on $\mathbb{N}_{k+1}$ fixing the last element. Let τ_l be the transposition $(l, k+1)$, for all $l = 1, ..., k, k+1$. Define $k+1$ permutations, $\sigma_{(l)}$, on $\mathbb{N}_{k+1}$, as follows:

$$(7) \qquad\qquad\qquad\qquad \sigma_{(l)} = \sigma\tau_l, \text{ for } l = 1, ..., k, k+1.$$

Informally speaking, given the cycle decomposition of σ, we obtain $\sigma_{(l)}$, for each $l = 1, ..., k$, by inserting the element $k+1$ immediately after the element l, and when $l = k+1$, the permutation $\sigma_{(k+1)}$ fixes the element $k+1$. Clearly $\sigma_{(l)}^{-1}(k+1) = l$ for all l, and

$$\{\text{All permutations on } \mathbb{N}_{k+1}\} = \{\sigma\tau_l \,|\, \sigma \text{ is a permutation on } \mathbb{N}_k, \ \ l = 1, ..., k, k+1\}.$$

For a fixed vector $\mu \in \mathbb{R}^n$ we define k linear maps

$$T \in T^{k,n} \rightarrow T_{\text{out}}^{(l)} \in T^{k+1,n}, \text{ for } l = 1, 2, ..., k,$$

as follows:

$$(8) \qquad \left(T_{\text{out}}^{(l)}\right)^{i_1 ... i_k i_{k+1}} = \begin{cases} 0, & \text{if } i_l \sim i_{k+1}, \\ \dfrac{T^{i_1 ... i_{l-1} i_{k+1} i_{l+1} ... i_k} - T^{i_1 ... i_{l-1} i_l i_{l+1} ... i_k}}{\mu_{i_{k+1}} - \mu_{i_l}}, & \text{if } i_l \not\sim i_{k+1}. \end{cases}$$

Notice that if T is a μ-block-constant tensor, then so is $T_{\text{out}}^{(l)}$ for each $l = 1, ..., k$. The next theorem is Corollary 5.8 from [9].

Theorem 2.4 *Let $\{M_m\}$ be a sequence of symmetric matrices converging to 0, such that $M_m/\|M_m\|$ converges to M. Let μ be in $\mathbb{R}_\downarrow^n$ and $U_m \rightarrow U \in O^n$ be a sequence of orthogonal matrices such that*

$$\text{Diag}\,\mu + M_m = U_m\big(\text{Diag}\,\lambda(\text{Diag}\,\mu + M_m)\big)U_m^T, \ \ \text{for all } m = 1, 2, ...$$

Then for every block-constant k-tensor T on $\mathbb{R}^n$, and any permutation σ on $\mathbb{N}_k$ we have

$$(9) \qquad \lim_{m\to\infty} \frac{U_m(\text{Diag}\,{}^\sigma T)U_m^T - \text{Diag}\,{}^\sigma T}{\|M_m\|} = \sum_{l=1}^{k} \big(\text{Diag}\,{}^{\sigma_{(l)}} T_{\text{out}}^{(l)}\big)[M].$$

Again, for a fixed vector $\mu \in \mathbb{R}^n$, we define k linear maps

$$T \in T^{k,n} \rightarrow T_{\text{in}}^{(l)} \in T^{k+1,n}, \text{ for } l = 1, 2, ..., k,$$

5

as follows:

$$(10) \qquad \left(T_{\text{in}}^{\tau_l}\right)^{i_1 \dots i_k i_{k+1}} = \begin{cases} T^{i_1 \dots i_{l-1} i_{k+1} i_{l+1} \dots i_k}, & \text{if } i_l \sim i_{k+1}, \\ 0, & \text{if } i_l \not\sim i_{k+1}. \end{cases}$$

Notice that if T is a block-constant tensor, then so is $T_{\text{in}}^{\tau_l}$ for each $l = 1, \dots, k$. Finally, we define

$$(11) \qquad \left(T^{\tau_l}\right)^{i_1 \dots i_k i_{k+1}} = \begin{cases} T^{i_1 \dots i_{l-1} i_l i_{l+1} \dots i_k}, & \text{if } i_l = i_{k+1}, \\ 0, & \text{if } i_l \neq i_{k+1}. \end{cases}$$

In other words, T^{τ_l} is a $(k+1)$-tensor with entries off the hyper plane $i_l = i_{k+1}$ equal to zero. On the hyper plane $i_l = i_{k+1}$ we have placed the original tensor T. The next theorem is Corollary 5.6 from [9].

Theorem 2.5 *Let $U \in O^n$ be a block-diagonal orthogonal matrix and let σ be a permutation on $\mathbb{N}_k$. Let M be an arbitrary symmetric matrix, and $h \in \mathbb{R}^n$ be a vector, such that $U^T M_{\text{in}} U = \text{Diag}\, h$. Then*

(i) for any block-constant $(k+1)$-tensor T on $\mathbb{R}^n$,

$$U\left(\text{Diag}^{\,\sigma}(T[h])\right)U^T = \left(\text{Diag}^{\,\sigma_{(k+1)}} T\right)[M];$$

(ii) for any block-constant k-tensor T on $\mathbb{R}^n$

$$U\left(\text{Diag}^{\,\sigma}(T^{\tau_l}[h])\right)U^T = \left(\text{Diag}^{\,\sigma_{(l)}} T_{\text{in}}^{\tau_l}\right)[M], \quad \text{for all } l = 1, \dots, k,$$

where the permutations $\sigma_{(l)}$, for $l \in \mathbb{N}_k$, are defined by (7).

3 Several standing assumptions

Our approach is to successively differentiate the composition $f \circ \lambda$ where at every step we use the tensorial language presented in Section 2 to simplify the calculation. More precisely, we will define k-tensor valued maps $\mathcal{A}_\sigma : \mathbb{R}^n \to T^{k,n}$, $\sigma \in P^k$, (only in terms of the function f and its partial derivatives) such that

$$(12) \qquad \nabla^k (f \circ \lambda)(X) = V\left(\sum_{\sigma \in P^k} \text{Diag}^{\,\sigma} \mathcal{A}_\sigma(\lambda(X))\right) V^T,$$

where $X = V(\text{Diag}\, \lambda(X))V^T$. The formula for the gradient (the case $k = 1$) was originally derived in [4], see also Subsection 5.1 below. We showed in [8, Section 5], that having derived that formula for $k = 1$, then for $k \geq 2$ it is enough to show it under the following three assumptions.

- The matrix X is diagonal, $\text{Diag}\, \mu$, for some vector $\mu \in \mathbb{R}^n_{\downarrow}$.
- The sequence $\{M_m\}$ of symmetric matrices converges to 0 and is such that $M_m/\|M_m\|$ converges to M.

- A sequence of orthogonal matrices $U_m \in O^n$ is chosen such that

$$\operatorname{Diag}\mu + M_m = U_m\big(\operatorname{Diag}\lambda(\operatorname{Diag}\mu + M_m)\big)U_m^T, \quad \text{for all} \ \ m = 1, 2, ...$$

and U_m approaches $U \in O^n$ as m goes to infinity. (U is block diagonal with blocks determined by μ.)

The next lemma (the proof is a simple combination of Lemma 5.10 in [5] and Theorem 3.12 in [3]) justifies the notation that follows. Recall that $\mu \in \mathbb{R}^n$ partitions $\mathbb{N}_n$ into r blocks $I_1, ..., I_r$.

Lemma 3.1 *For any $\mu \in \mathbb{R}^n_\downarrow$ and sequence of symmetric matrices $M_m \to 0$ we have that*

$$(13) \qquad \lambda(\operatorname{Diag}\mu + M_m)^T = \mu^T + \big(\lambda(X_1^T M_m X_1)^T, ..., \lambda(X_r^T M_m X_r)^T\big)^T + o(\|M_m\|),$$

where $X_l := [e^i \,|\, i \in I_l]$, for all $l = 1, ..., r$.

Throughout the whole paper, we denote

$$(14) \qquad\qquad h_m := \big(\lambda(X_1^T M_m X_1)^T, ..., \lambda(X_r^T M_m X_r)^T\big)^T .$$

If also $M_m/\|M_m\|$ converges to M as m goes to infinity, since the eigenvalues are continuous functions, we can define

$$(15) \qquad\qquad h := \lim_{m \to \infty} \frac{h_m}{\|M_m\|} = \big(\lambda(X_1^T M X_1)^T, ..., \lambda(X_r^T M X_r)^T\big)^T .$$

We reserve the symbols h_m and h to denote the above two vectors throughout the paper. With this notation Lemma 3.1 says that if $M_m \to 0$, then

$$(16) \qquad\qquad \lambda(\operatorname{Diag}\mu + M_m)^T = \mu^T + h_m + o(\|M_m\|).$$

If, for the fixed vector $\mu \in \mathbb{R}^n_\downarrow$, we define

$$M_{\text{in}}^{ij} = \begin{cases} M^{ij}, & \text{if } i \sim j, \\ 0, & \text{otherwise,} \end{cases}$$

then Theorem 4.2 in [9] says that the orthogonal matrix U is block-diagonal and satisfies

$$(17) \qquad\qquad U^T M_{\text{in}} U = \operatorname{Diag} h.$$

4   Analyticity of isolated eigenvalues

Let A be in S^n and suppose that the j-th largest eigenvalue is isolated, that is

$$\lambda_{j-1}(A) > \lambda_j(A) > \lambda_{j+1}(A).$$

The goal of this section is to give two justifications of the known fact that $\lambda_j(\cdot)$ is an analytic function in a neighbourhood of A. We call a function of several real variables *analytic* at a point if in a neighbourhood of this point it has an power series expansion. The corresponding complex variable notion is called *holomorphic*.

The first justification below is from [11, Theorem 2.1].

Theorem 4.1 *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a function analytic at the point $\lambda(A)$ for some A in S^n. Suppose also $f(Px) = f(x)$ for every permutation matrix, P, for which $P\lambda(A) = \lambda(A)$. Then, the function $f \circ \lambda$ is analytic at A.* ∎

To see how this theorem implies the analyticity of $\lambda_j(\cdot)$ take

$$f(x_1, ..., x_n) = \text{the } j^{\text{th}} \text{ largest element of } \{x_1, ..., x_n\}.$$

The function f is a piece-wise affine function. Moreover, for any $x \in \mathbb{R}^n$ in a neighbourhood of the vector $\lambda(A)$ it is given by

$$f(x) = x_j.$$

Thus, f is analytic in that neighbourhood. Next, f is a symmetric function and thus by definition $f(Px) = f(x)$ for every $x \in R^n$ and every permutation matrix P. Therefore by the theorem $\lambda_j = f \circ \lambda$ is an analytic function.

For the second justification we use the following result [1]. (In the theorem below, $\lambda_i(X)$ denotes an arbitrary eigenvalue of a matrix X, not necessarily the i'th largest one.)

Theorem 4.2 (Arnold 1971) *Suppose that $A \in \mathbb{C}^{n \times n}$ has q eigenvalues $\lambda_1(A), ..., \lambda_q(A)$ (counting multiplicities) in an open set $\Omega \subset \mathbb{C}$, and the rest $n - q$ eigenvalues not in the closure of Ω. Then, there is a neighbourhood Δ of A and holomorphic mappings $S : \Delta \to \mathbb{C}^{q \times q}$ and $T : \Delta \to \mathbb{C}^{(n-q) \times (n-q)}$ such that for all $X \in \Delta$*

$$X \text{ is similar to } \begin{pmatrix} S(X) & 0 \\ 0 & T(X) \end{pmatrix},$$

and $S(A)$ has eigenvalues $\lambda_1(A), ..., \lambda_q(A)$. ∎

To deduce the result we need, since the jth largest eigenvalue is isolated, we can find an open set $\Omega \subset \mathbb{C}$, such that only that eigenvalue is in Ω and the remaining $n - 1$ are not in the closure of Ω. By the theorem, there is a neighbourhood Δ of A and holomorphic mapping $S : \Delta \to \mathbb{C}$ such that $S(X)$ is equal to the jth largest eigenvalue of X for all X in Δ.

If A is a real symmetric matrix, then the intersection of Δ with S^n is a neighbourhood of A in S^n. Let $\tilde{S}(X)$ denote the restriction of $S(X)$ to $\Delta \cap S^n$. Clearly, $\tilde{S}(X)$ is holomorphic, real valued function. Therefore, (it is a standard result in complex analysis) the coefficients in the power series expansion of $\tilde{S}(X)$ must be real numbers. Thus, the jth largest eigenvalue is a real analytic function in the neighbourhood $\Delta \cap S^n$ or A.

All these considerations make the following observation clear.

Theorem 4.3 *Suppose that $A \in S^n$ has distinct eigenvalues and $f : \mathbb{R}^n \to \mathbb{R}$ is k-times (continuously) differentiable in a neighbourhood of $\lambda(A)$. Then, $f \circ \lambda$ is k-times (continuously) differentiable in a neighbourhood of A.*

5   The kth derivative of functions of eigenvalues at a matrix with distinct eigenvalues

Let $f : \mathbb{R}^n \to \mathbb{R}$ be an arbitrary k-times (continuously) differentiable function. In this section, we do not assume that f is a symmetric function. Our goal in this section is to derive a formula for the kth derivative of $f \circ \lambda$ on the set $\lambda^{-1}(\Omega)$, where

$$\Omega = \{x \in \mathbb{R}^n \mid x_i \neq x_j \text{ for every } i \neq j\}, \text{ and}$$
$$\lambda^{-1}(\Omega) = \{A \in S^n \mid \lambda(A) \in \Omega\}.$$

Clearly Ω is a dense, open subset of $\mathbb{R}^n$ and $\lambda^{-1}(\Omega)$ is a dense, open subset of S^n.

As an example of how one can differentiate $f \circ \lambda$, let us consider the general situation. Let X, Y, and Z be Banach spaces and let $g : X \to Y$, $G : Y \to Z$. Then, by applying the chain rule we have the following formulae for the first three derivatives of $\phi = G \circ g$, (see [2, Section X.4]) for any vectors h_1, h_2, h_3 from X:

$$\nabla\phi(x)[h_1] = \nabla G(g(x))[\nabla g(x)[h_1]],$$
$$\nabla^2\phi(x)[h_1, h_2] = \nabla^2 G(g(x))[\nabla g(x)[h_1], \nabla g(x)[h_2]] + \nabla G(g(x))[\nabla^2 g(x)[h_1, h_2]],$$
$$\begin{aligned}
\nabla^3\phi(x)[h_1, h_2, h_3] = {} & \nabla^3 G(g(x))[\nabla g(x)[h_1], \nabla g(x)[h_2], \nabla g(x)[h_3]] \\
& + \nabla^2 G(g(x))[\nabla g(x)[h_1], \nabla^2 g(x)[h_2, h_3]] \\
& + \nabla^2 G(g(x))[\nabla g(x)[h_2], \nabla^2 g(x)[h_1, h_3]] \\
& + \nabla^2 G(g(x))[\nabla g(x)[h_3], \nabla^2 g(x)[h_1, h_2]] \\
& + \nabla G(g(x))[\nabla^3 g(x)[h_1, h_2, h_3]].
\end{aligned}$$

In our case, we have $X = S^n$, $Y = \mathbb{R}^n$, $Z = \mathbb{R}$, $g = \lambda$, and $G = f$. As can be seen from the above example, this approach very quickly becomes unmanageable. The formula for the k-derivative of the composition requires formulae for every derivative of λ up to the kth. It is not clear how one can organize and simplify the resulting expression into a compact, ordered formula.

Fix a vector $\mu \in \mathbb{R}^n_{\downarrow} \cap \Omega$. Since μ has distinct entries, every block in the partition that it defines will have exactly one element. This means that for any $j, i \in \mathbb{N}_n$, $i \sim j \Leftrightarrow i = j$, and that makes any tensor block-constant. In particular for the matrices X_l, defined in Lemma 3.1, we have $X_l = [e^l]$, $l = 1, ..., n$. This implies that $h_m = \operatorname{diag} M_m$ and that $h = \operatorname{diag} M$. Notice, finally, how the definition of $T_{\text{out}}^{(l)}$ changes:

$$(18) \qquad \left(T_{\text{out}}^{(l)}\right)^{i_1 \dots i_k i_{k+1}} = \begin{cases} 0, & \text{if } i_l = i_{k+1}, \\ \frac{T^{i_1 \dots i_{l-1} i_{k+1} i_{l+1} \dots i_k} - T^{i_1 \dots i_{l-1} i_l i_{l+1} \dots i_k}}{\mu_{i_{k+1}} - \mu_{i_l}}, & \text{if } i_l \neq i_{k+1}. \end{cases}$$

We will derive Formula (12) by induction on the order of the derivative. For completeness, we begin by recalculating the formula for the gradient.

5.1 The gradient

Using Formulae (16) we compute

$$\lim_{m\to\infty} \frac{(f\circ\lambda)(\operatorname{Diag}\mu + M_m) - (f\circ\lambda)(\operatorname{Diag}\mu)}{\|M_m\|} = \lim_{m\to\infty} \frac{f(\mu + h_m + o(\|M_m\|)) - f(\mu)}{\|M_m\|}$$

$$= \lim_{m\to\infty} \frac{f(\mu) + \nabla f(\mu)[h_m] + o(\|M_m\|) - f(\mu)}{\|M_m\|}$$

$$= \nabla f(\mu)[h]$$

$$= \langle \nabla f(\mu), \operatorname{diag} M \rangle$$

$$= \big(\operatorname{Diag}\nabla f(\mu)\big)[M].$$

This shows that $\nabla(f\circ\lambda)(\operatorname{Diag}\mu) = \operatorname{Diag}^{(1)}\nabla f(\mu)$. It is easy to see now that

$$(19) \qquad \nabla(f\circ\lambda)(X) = V\big(\operatorname{Diag}^{(1)}\nabla f(\lambda(X))\big)V^T = V\Big(\sum_{\sigma\in P^1}\operatorname{Diag}^\sigma \mathcal{A}_\sigma(\lambda(X))\Big)V^T,$$

where $X = V(\operatorname{Diag}\lambda(X))V^T$ and $\mathcal{A}_{(1)}(x) = \nabla f(x)$. Trivially, if f is k-times (continuously) differentiable, then $\mathcal{A}_{(1)}(x) = \nabla f(x)$ is $(k-1)$-times (continuously) differentiable.

Note that when the eigenvalues of X are not distinct, the calculation of the gradient of $f\circ\lambda$ is almost identical and leads to the same final formula. Indeed, using Equation (17),

$$\nabla f(\mu)[h] = \langle \nabla f(\mu), \operatorname{diag}(U^T M_{\mathrm{in}} U)\rangle = \big(U(\operatorname{Diag}\nabla f(\mu))U^T\big)[M] = \big(\operatorname{Diag}\nabla f(\mu)\big)[M],$$

where in the last equality we used the fact the U is block-diagonal, orthogonal and $\nabla f(\mu)$ is block-constant.

5.2 The induction step

Suppose now that for some $1 \le s < k$

$$\nabla^s(f\circ\lambda)(X) = V\Big(\sum_{\sigma\in P^s}\operatorname{Diag}^\sigma \mathcal{A}_\sigma(\lambda(X))\Big)V^T,$$

where $X = V(\operatorname{Diag}\lambda(X))V^T$. Suppose also that for every $\sigma \in P^s$, the s-tensor valued map $\mathcal{A}_\sigma : \mathbb{R}^n \to T^{s,n}$, is $(k-s)$-times (continuously) differentiable.

Using Formulae (16), we differentiate $\nabla^s(f\circ\lambda)$ at the matrix $\operatorname{Diag}\mu$:

$$\nabla^{s+1}(f\circ\lambda)(\operatorname{Diag}\mu)[M]$$

$$= \lim_{m\to\infty} \frac{\nabla^s(f\circ\lambda)(\operatorname{Diag}\mu + M_m) - \nabla^s(f\circ\lambda)(\operatorname{Diag}\mu)}{\|M_m\|}$$

$$= \lim_{m\to\infty} \frac{U_m\big(\sum_{\sigma\in P^s}\operatorname{Diag}^\sigma \mathcal{A}_\sigma(\lambda(\operatorname{Diag}\mu + M_m))\big)U_m^T - \sum_{\sigma\in P^s}\operatorname{Diag}^\sigma \mathcal{A}_\sigma(\mu)}{\|M_m\|}$$

$$= \lim_{m \to \infty} \frac{\sum_{\sigma \in P^s} \Big(U_m \big(\operatorname{Diag}^{\sigma} \mathcal{A}_{\sigma}(\lambda(\operatorname{Diag}\mu + M_m))) \big) U_m^T - \operatorname{Diag}^{\sigma} \mathcal{A}_{\sigma}(\mu) \Big)}{\|M_m\|}$$

$$= \lim_{m \to \infty} \frac{\sum_{\sigma \in P^s} \Big(U_m \big(\operatorname{Diag}^{\sigma} \mathcal{A}_{\sigma}(\mu + h_m + o(\|M_m\|)) \big) U_m^T - \operatorname{Diag}^{\sigma} \mathcal{A}_{\sigma}(\mu) \Big)}{\|M_m\|}$$

$$= \lim_{m \to \infty} \frac{\sum_{\sigma \in P^s} \Big(U_m \big(\operatorname{Diag}^{\sigma} \big(\mathcal{A}_{\sigma}(\mu) + \nabla \mathcal{A}_{\sigma}(\mu)[h_m] + o(\|M_m\|) \big) \big) U_m^T - \operatorname{Diag}^{\sigma} \mathcal{A}_{\sigma}(\mu) \Big)}{\|M_m\|}$$

$$= \lim_{m \to \infty} \sum_{\sigma \in P^s} \frac{U_m \big(\operatorname{Diag}^{\sigma} \mathcal{A}_{\sigma}(\mu) \big) U_m^T - \operatorname{Diag}^{\sigma} \mathcal{A}_{\sigma}(\mu)}{\|M_m\|} + \sum_{\sigma \in P^s} U \big(\operatorname{Diag}^{\sigma} \big(\nabla \mathcal{A}_{\sigma}(\mu)[h] \big) \big) U^T.$$

By Theorem 2.4, since for every $\sigma \in P^s$, the tensor $\mathcal{A}_{\sigma}(\mu)$ is block-constant, we have

$$\lim_{m \to \infty} \frac{U_m \big(\operatorname{Diag}^{\sigma} \mathcal{A}_{\sigma}(\mu) \big) U_m^T - \operatorname{Diag}^{\sigma} \mathcal{A}_{\sigma}(\mu)}{\|M_m\|} = \sum_{l=1}^{s} \big(\operatorname{Diag}^{\sigma_{(l)}} (\mathcal{A}_{\sigma}(\mu))_{\text{out}}^{(l)} \big)[M].$$

By Theorem 2.5, since for every $\sigma \in P^s$ $\nabla \mathcal{A}_{\sigma}(\mu)$ is a block-constant $(s+1)$-tensor, we have

$$U \big(\operatorname{Diag}^{\sigma} \big(\nabla \mathcal{A}_{\sigma}(\mu)[h] \big) \big) U^T = \big(\operatorname{Diag}^{\sigma_{(s+1)}} \nabla \mathcal{A}_{\sigma}(\mu) \big)[M].$$

Thus we define

$$\mathcal{A}_{\sigma_{(l)}} := (\mathcal{A}_{\sigma}(\mu))_{\text{out}}^{(l)}, \quad \text{for all } l \in \mathbb{N}_s, \text{ and}$$
$$\mathcal{A}_{\sigma_{(s+1)}} := \nabla \mathcal{A}_{\sigma}(\mu).$$

Putting everything together and conclude that for every symmetric matrix M:

$$\nabla^{s+1}(f \circ \lambda)(\operatorname{Diag}\mu)[M] = \Big(\sum_{\substack{\sigma \in P^s \\ l \in \mathbb{N}_{s+1}}} \operatorname{Diag}^{\sigma_{(l)}} \mathcal{A}_{\sigma_{(l)}}(\mu) \Big)[M].$$

Notice the parameters of the summation sign in the above formula. As σ goes over the elements of P^s and as l goes over the set $\mathbb{N}_{s+1}$ the permutation $\sigma_{(l)}$ covers, in a one-to-one manner, all permutations in P^{s+1}. Now, the comments in [8, Section 5] show that

$$\nabla^{s+1}(f \circ \lambda)(X) = V \Big(\sum_{\substack{\sigma \in P^s \\ l \in \mathbb{N}_{s+1}}} \operatorname{Diag}^{\sigma_{(l)}} \mathcal{A}_{\sigma_{(l)}}(\lambda(X)) \Big) V^T,$$

where $X = V(\operatorname{Diag}\lambda(X))V^T$.

To finish the induction, we have to show that the $(s+1)$-tensor valued maps $\mathcal{A}_{\sigma_{(l)}}(\cdot)$ are at least $(k - s - 1)$-times (continuously) differentiable. This is clear when $l = s+1$ and $\sigma \in P^s$, since $\mathcal{A}_{\sigma}(\cdot)$ is $(k - s)$-times (continuously) differentiable for every $\sigma \in P^s$. For the rest of the maps this

11

is also easy to see. Every entry in $\mathcal{A}_{\sigma_{(l)}}$ is the difference of two entries of $\mathcal{A}_\sigma$ divided by a quantity that never becomes zero over the set Ω. This shows that over the set Ω, $\mathcal{A}_{\sigma_{(l)}}(\cdot)$ is $(k-s)$-times (continuously) differentiable for every $\sigma \in P^s$ and every $l \in \mathbb{N}_s$.

We summarize everything in the next theorem.

Theorem 5.1 *Let X be a symmetric matrix with distinct eigenvalues and let f be a k-times (continuously) differentiable function on $\mathbb{R}^n$. Let V be an orthogonal matrix such that $X = V(\mathrm{Diag}\,\lambda(X))V^T$. Then, $f \circ \lambda$ is k-times (continuously) differentiable function at X. Moreover if $\nabla^s(f \circ \lambda)$, for some $s < k$, is given by*

$$\nabla^s(f \circ \lambda)(X) = V\Big(\sum_{\sigma \in P^s} \mathrm{Diag}^\sigma \mathcal{A}_\sigma(\lambda(X))\Big)V^T,$$

for some s-tensor valued mappings $\mathcal{A}_\sigma : \mathbb{R}^n \to T^{s,n}$, for every $\sigma \in P^s$, then $\nabla^{(s+1)}(f \circ \lambda)$ is given by

$$(20) \qquad \nabla^{(s+1)}(f \circ \lambda)(X) = V\Big(\sum_{\substack{\sigma \in P^s \\ l \in \mathbb{N}_{s+1}}} \mathrm{Diag}^{\sigma_{(l)}} \mathcal{A}_{\sigma_{(l)}}(\lambda(X))\Big)V^T,$$

where

$$\mathcal{A}_{\sigma_{(l)}} = \big(\mathcal{A}_\sigma\big)^{(l)}_{\mathrm{out}}, \;\; \text{for all } l \in \mathbb{N}_s, \;\; \text{and}$$
$$\mathcal{A}_{\sigma_{(s+1)}} = \nabla\mathcal{A}_\sigma.$$

6 The kth derivative of separable spectral functions

In this section we show that Formula (12) holds at an arbitrary symmetric matrix X (not necessarily with distinct eigenvalues) for a subclass of spectral functions that we now describe.

Let g be a real function on an interval I. If $D = \mathrm{Diag}\,(\lambda_1, ..., \lambda_n)$ is a diagonal matrix with diagonal entries $\lambda_i \in I$, $i = 1, ..., n$, we define

$$(21) \qquad\qquad G(D) = \mathrm{Diag}\,(g(\lambda_1), ..., g(\lambda_1)).$$

If X is a symmetric matrix with eigenvalues λ_i in I, we choose an orthogonal matrix V such that $X = V(\mathrm{Diag}\,\lambda(X))V^T$ and, then define

$$(22) \qquad\qquad G(X) = VG(\mathrm{Diag}\,\lambda(X))V^T.$$

In this way we obtain a (well-defined) symmetric-matrix valued function with domain the set of all matrices X with eigenvalues in I.

These functions have been the object of recent interest in optimization [10] and the main object of [2, Chapter V], where their gradient is computed using an approximation argument. Notice that $G(X)$ is just the gradient (see Formula (19)) of the spectral function $f \circ \lambda$, where $f(x) = \tilde{g}(x_1) + \cdots + \tilde{g}(x_n)$, and $\tilde{g}(s) = \int_0^s g(t)\,dt$. That is why we will call those functions *separable spectral functions*.

6.1 Description of the kth derivative

Let $g : I \to \mathbb{R}$ be k-times (continuously) differentiable. Define the symmetric function $g^{[(12)]}(x,y) : I \times I \to \mathbb{R}$ as

(23)
$$g^{[(12)]}(x,y) = \begin{cases} \dfrac{g(x) - g(y)}{x - y}, & \text{if } x \neq y, \\ g'(x), & \text{if } x = y. \end{cases}$$

The integral representation $g^{[(12)]}(x,y) = \int_0^1 g'(y + t(x - y)) \, dt$ shows that $g^{[(12)]}(x,y)$ is as smooth, in both arguments, as g'.

Denote by $\tilde{P}^k$ the set of all permutations from P^k that have one cycle in their cycle decomposition. Clearly $|\tilde{P}^k| = (k-1)!$. Notice that for every $\sigma \in \tilde{P}^k$ and every $l \in \mathbb{N}_k$ we have $\sigma_{(l)} \in \tilde{P}^{k+1}$. Moreover, as σ varies over $\tilde{P}^k$ and l varies over $\mathbb{N}_k$, the permutation $\sigma_{(l)}$ varies over $\tilde{P}^{k+1}$ in a one-to-one and onto fashion.

Suppose that for every $\sigma \in \tilde{P}^k$ we have defined the function $g^{[\sigma]}(x_1, ..., x_k)$ on the set $I \times I \times \cdots \times I$, k-times, and suppose that these functions are as smooth as $g^{(k-1)}$ (the $(k-1)$-th derivative of g). For every $\sigma \in \tilde{P}^k$ and every $l \in \mathbb{N}_k$ we define the function $g^{[\sigma_{(l)}]}(x_1, ..., x_k, x_{k+1})$ as follows:

(24)
$$g^{[\sigma_{(l)}]}(x_1, ..., x_{k+1}) = \begin{cases} \nabla_l g^{[\sigma]}(x_1, ..., x_k), & \text{if } x_l = x_{k+1} \\ \dfrac{g^{[\sigma]}(x_1, ..., x_l, ..., x_k) - g^{[\sigma]}(x_1, ..., x_{k+1}, ..., x_k)}{x_l - x_{k+1}}, & \text{if } x_l \neq x_{k+1}, \end{cases}$$

where in the second case of the definition, both x_l and x_{k+1} are in l-th position, and ∇_l denotes the partial derivative with respect to the l-th argument. Using the integral formula

$$g^{[\sigma_{(l)}]}(x_1, ..., x_{k+1}) = \int_0^1 \nabla_l g^{[\sigma]}(x_1, ..., x_{l-1}, x_{k+1} + t(x_l - x_{k+1}), x_{l+1}, ..., x_k) \, dt,$$

for every $l \in \mathbb{N}_k$, we see that $g^{[\sigma_{(l)}]}(x_1, ..., x_{k+1})$ is as smooth as $g^{(k)}$, the k-th derivative of g.

Finally, for every $s \in \{2, 3, ..., k+1\}$ and every $\sigma \in \tilde{P}^s$, we define a s-tensor valued map

(25)
$$\mathbf{g}^{[\sigma]} : \mathbb{R}^n \to T^{s,n}, \text{ where}$$
$$\left(\mathbf{g}^{[\sigma]}(\mu)\right)^{i_1 ... i_s} := g^{[\sigma]}(\mu_{i_1}, ..., \mu_{i_s}).$$

Clearly, if $(i_1, ..., i_s) \sim_\mu (j_1, ..., j_s)$, then $\left(\mathbf{g}^{[\sigma]}(\mu)\right)^{i_1 ... i_s} = \left(\mathbf{g}^{[\sigma]}(\mu)\right)^{j_1 ... j_s}$, which shows that $\mathbf{g}^{[\sigma]}(\mu)$ is a μ-block-constant tensor for every μ. Moreover, the map $\mathbf{g}^{[\sigma]} : \mathbb{R}^n \to T^{s,n}$ is still as smooth as $g^{(s-1)}$, for every $s = 2, 3, ..., k+1$.

We are now ready to formulate the second main result of this work. The proof is given in the next subsection. A comparison between Theorem 6.1 and Theorem 5.1 is given at the end of Subsection 6.3.

Theorem 6.1 *Let g be a k-times (continuously) differentiable function defined on an interval I. Let X be a symmetric matrix with eigenvalues in the interval I, and let V be an orthogonal matrix such that $X = V(\operatorname{Diag} \lambda(X))V^T$. Then, the matrix valued function G defined by (21) and (22) is k-times (continuously) differentiable at X. Moreover its k-th derivative, $\nabla^k G(X)$, is given by the formula*

$$(26) \qquad \nabla^k G(X) = V\Big(\sum_{\sigma \in \tilde{P}^{k+1}} \operatorname{Diag}^\sigma \mathbf{g}^{[\sigma]}(\lambda(X)) \Big) V^T,$$

where the $(k+1)$-tensor valued maps $\mathbf{g}^{[\sigma]}(\cdot)$ are defined by Equation (25).

6.2   Proof of Theorem 6.1: the gradient

Let X be an $n \times n$ symmetric matrix with all eigenvalues in I and such that $X = V(\operatorname{Diag} \lambda(X))V^T$ for some orthogonal matrix V. The formula for the gradient of separable spectral functions has been known for a while. For example, using approximation techniques, it was shown in [2] that for any two symmetric matrices H_1 and H_2

$$(27) \qquad \nabla G(X)[H_1, H_2] = \langle V\big(g^{[(12)]}(\lambda(X)) \circ (V^T H_1 V)\big)V^T, H_2 \rangle,$$

where '$\circ$' stands for the usual Hadamard product.

In this subsection, we will give a direct derivation of the gradient and as a result a slightly different representation of the above formula.

For convenience we denote $g(x) := (g(x_1), ..., g(x_n))^T$, and $\nabla g(x) := (g'(x_1), ..., g'(x_n))^T$, for any $x \in \mathbb{R}^n$. Thus we compute:

$$
\begin{aligned}
\nabla G(\operatorname{Diag} \mu)[M] &= \lim_{m\to\infty} \frac{G(\operatorname{Diag} \mu + M_m) - G(\operatorname{Diag} \mu)}{\|M_m\|} \\
&= \lim_{m\to\infty} \frac{U_m \operatorname{Diag} g(\lambda(\operatorname{Diag} \mu + M_m)) U_m^T - \operatorname{Diag} g(\mu)}{\|M_m\|} \\
&= \lim_{m\to\infty} \frac{U_m (\operatorname{Diag} g(\mu + h_m + o(\|M_m\|))) U_m^T - \operatorname{Diag} g(\mu)}{\|M_m\|} \\
&= \lim_{m\to\infty} \frac{U_m \big(\operatorname{Diag} g(\mu) + (\operatorname{Diag} \nabla g(\mu))[h_m] + o(\|M_m\|)\big) U_m^T - \operatorname{Diag} g(\mu)}{\|M_m\|} \\
&= \lim_{m\to\infty} \frac{U_m \big(\operatorname{Diag} g(\mu)\big) U_m^T - \operatorname{Diag} g(\mu)}{\|M_m\|} + U\big((\operatorname{Diag} \nabla g(\mu))[h]\big) U^T.
\end{aligned}
$$

It is important to notice that both vectors $g(\mu)$ and $\nabla g(\mu)$ are block-constant (with respect to μ). We use the second part of Corollary 2.5 with $\sigma = (1)$, $k = l = 1$, (notice that $\sigma_{(l)} = (12)$) and $T = \nabla g(\mu)$ to develop the second term above:

$$U\big((\operatorname{Diag} \nabla g(\mu))[h]\big) U^T = \big(\operatorname{Diag}^{(12)}(\nabla g(\mu))_{\text{in}}^{(12)}\big)[M].$$

14

We now use Corollary 2.4 with $k = 1$, $\sigma = (1)$, and $T = g(\mu)$ to find the limit:

$$\lim_{m \to \infty} \frac{U_m\big(\mathrm{Diag}\, g(\mu)\big)U_m^T - \mathrm{Diag}\, g(\mu)}{\|M_m\|} = \big(\mathrm{Diag}^{\,(12)}(g(\mu))_{\mathrm{out}}^{(12)}\big)[M].$$

Putting everything together we get

$$\nabla G(\mathrm{Diag}\,\mu) = \mathrm{Diag}^{\,(12)}(\nabla g(\mu))_{\mathrm{in}}^{(12)} + \mathrm{Diag}^{\,(12)}(g(\mu))_{\mathrm{out}}^{(12)} = \mathrm{Diag}^{\,(12)}\mathbf{g}^{[(12)]}(\mu),$$

where we used the easy to check fact that $\mathbf{g}^{[(12)]}(\mu) = (\nabla g(\mu))_{\mathrm{in}}^{(12)} + (g(\mu))_{\mathrm{out}}^{(12)}$. Now, using Lemma 2.3 it is easy to see the following result (when X is arbitrary symmetric matrix, not just diagonal).

Theorem 6.2 *Let $g \in C^1(I)$ and let X be a symmetric matrix with all eigenvalues in I. Then,*

$$(28) \qquad\qquad \nabla G(X) = V\big(\mathrm{Diag}^{\,(12)}\mathbf{g}^{[(12)]}(\lambda(X))\big)V^T,$$

where $X = V(\mathrm{Diag}\,\lambda(X))V^T$.

For the sake of completeness, we show that Formula (28) is indeed the same as Formula (27). This is achieved when in the next result one substitutes the matrix A with the matrix $\mathbf{g}^{[(12)]}(\lambda(X))$.

Proposition 6.3 *For any $n \times n$ matrix A, any orthogonal V, and any symmetric H_1 and H_2, we have the equality*

$$\big(V(\mathrm{Diag}^{\,(12)}A)V^T\big)[H_1, H_2] = \langle V\big(A \circ (V^T H_1 V)\big)V^T, H_2\rangle,$$

where '$\circ$' stands for the ordinary Hadamard product.

Proof. We develop the two sides of the stated equality and compare the results. By Theorem 2.2, the left-hand side is equal to

$$V\big(\mathrm{Diag}^{\,(12)}A\big)V^T[H_1, H_2] = \langle A, \tilde{H}_1 \circ_{(12)} \tilde{H}_2\rangle.$$

On the other hand

$$\langle V\big(A \circ (V^T H_1 V)\big)V^T, H_2\rangle = \langle A \circ \tilde{H}_1, \tilde{H}_2\rangle = \langle A, \tilde{H}_1 \circ \tilde{H}_2\rangle.$$

Finally it is easy to check directly from the definitions that $\tilde{H}_1 \circ_{(12)} \tilde{H}_2 = \tilde{H}_1 \circ \tilde{H}_2^T = \tilde{H}_1 \circ \tilde{H}_2$, where in the last equality we used that $\tilde{H}_2$ is symmetric. ∎

15

6.3 Proof of Theorem 6.1: the induction step

Suppose that $g : I \to \mathbb{R}$ is k-times (continuously) differentiable, and that the formula for the $(s-1)$-th derivative $(2 \le s < k+1)$ of G at the matrix X is given by

$$\nabla^{(s-1)}G(X) = V\Big(\sum_{\sigma \in \tilde{P}^s} \mathrm{Diag}^{\,\sigma} \mathbf{g}^{[\sigma]}(\lambda(X)) \Big) V^T.$$

The s-tensor-valued maps $\mathbf{g}^{[\sigma]} : \mathbb{R} \to T^{s,n}$ are at least $(k-s+1)$-times (continuously) differentiable. As we explained in Section 3, it is enough to derive the formula for $\nabla^s G(X)$ only in the case when $X = \mathrm{Diag}\,\mu$ for some $\mu \in \mathbb{R}^n_\downarrow$. We compute:

$$\nabla^s G(\mathrm{Diag}\,\mu)[M] = \lim_{m\to\infty} \frac{\nabla^{(s-1)}G(\mathrm{Diag}\,\mu + M_m) - \nabla^{(s-1)}G(\mathrm{Diag}\,\mu)}{\|M_m\|}$$

$$= \lim_{m\to\infty} \frac{U_m\Big(\sum_{\sigma\in\tilde{P}^s} \mathrm{Diag}^{\,\sigma} \mathbf{g}^{[\sigma]}(\lambda(\mathrm{Diag}\,\mu + M_m)) \Big) U_m^T - \sum_{\sigma\in\tilde{P}^s} \mathrm{Diag}^{\,\sigma} \mathbf{g}^{[\sigma]}(\mu)}{\|M_m\|}$$

$$= \lim_{m\to\infty} \frac{U_m\Big(\sum_{\sigma\in\tilde{P}^s} \mathrm{Diag}^{\,\sigma} \mathbf{g}^{[\sigma]}(\mu + h_m + o(\|M_m\|)) \Big) U_m^T - \sum_{\sigma\in\tilde{P}^s} \mathrm{Diag}^{\,\sigma} \mathbf{g}^{[\sigma]}(\mu)}{\|M_m\|}$$

$$= \lim_{m\to\infty} \frac{U_m\Big(\sum_{\sigma\in\tilde{P}^s} \mathrm{Diag}^{\,\sigma} \big(\mathbf{g}^{[\sigma]}(\mu) + \nabla\mathbf{g}^{[\sigma]}(\mu)[h_m] + o(\|M_m\|) \big) \Big) U_m^T - \sum_{\sigma\in\tilde{P}^k} \mathrm{Diag}^{\,\sigma} \mathbf{g}^{[\sigma]}(\mu)}{\|M_m\|}$$

$$= \lim_{m\to\infty} \frac{U_m\Big(\sum_{\sigma\in\tilde{P}^s} \mathrm{Diag}^{\,\sigma} \mathbf{g}^{[\sigma]}(\mu) \Big) U_m^T - \sum_{\sigma\in\tilde{P}^s} \mathrm{Diag}^{\,\sigma} \mathbf{g}^{[\sigma]}(\mu)}{\|M_m\|} + U\Big(\sum_{\sigma\in\tilde{P}^s} \mathrm{Diag}^{\,\sigma} \big(\nabla\mathbf{g}^{[\sigma]}(\mu)[h] \big) \Big) U^T.$$

First, using Theorem 2.4, we wrap up the limit in the above formula:

$$(29) \quad \lim_{m\to\infty} \frac{U_m\Big(\sum_{\sigma\in\tilde{P}^s} \mathrm{Diag}^{\,\sigma} \mathbf{g}^{[\sigma]}(\mu) \Big) U_m^T - \sum_{\sigma\in\tilde{P}^s} \mathrm{Diag}^{\,\sigma} \mathbf{g}^{[\sigma]}(\mu)}{\|M_m\|} = \sum_{\substack{\sigma\in\tilde{P}^s \\ l\in\mathbb{N}_s}} \big(\mathrm{Diag}^{\,\sigma_{(l)}} (\mathbf{g}^{[\sigma]}(\mu))^{(l)}_{\mathrm{out}} \big)[M].$$

Next, we focus our attention on the gradient $\nabla\mathbf{g}^{[\sigma]}(\mu)$. Using the definition, Equation (25), we see that

$$(30) \qquad \begin{aligned} \nabla\big[(\mathbf{g}^{[\sigma]}(\mu))^{i_1\ldots i_s}\big] &= \sum_{l=1}^s \nabla_l g^{[\sigma]}(\mu_{i_1}, \ldots, \mu_{i_s}) e^{i_l} \\ &= \sum_{l=1}^s g^{[\sigma_{(l)}]}(\mu_{i_1}, \ldots, \mu_{i_s}, \mu_{i_l}) e^{i_l}, \end{aligned}$$

where for the second equality we used Equation (24). This prompts us to define the s-tensor-valued map

$$T_l : \mathbb{R}^n \to T^{s,n}, \quad \text{where}$$

16

(31)
$$(T_l(\mu))^{i_1...i_s} := g^{[\sigma_{(l)}]}(\mu_{i_1}, ..., \mu_{i_s}, \mu_{i_l}).$$

for every $l \in \mathbb{N}_s$. Notice that $T_l(\mu)$ is a μ-block-constant s-tensor, for every μ and every $l \in \mathbb{N}_s$.

Lemma 6.4 *The gradient of* $\mathbf{g}^{[\sigma]}(\mu)$ *allows the following decomposition*

(32)
$$\nabla \mathbf{g}^{[\sigma]}(\mu) = \sum_{l=1}^{s} (T_l(\mu))^{\tau_l},$$

where the "lifting" $(T_l(\mu))^{\tau_l}$ *is defined by Equation (11).*

Proof. Fix a multi index $(i_1, ..., i_s)$. By definition of the gradient $\nabla g^{[\sigma]}(\mu)$ we have that

$$\left((\nabla \mathbf{g}^{[\sigma]}(\mu))^{i_1...i_s,1}, (\nabla \mathbf{g}^{[\sigma]}(\mu))^{i_1...i_s,2}, ..., (\nabla \mathbf{g}^{[\sigma]}(\mu))^{i_1...i_s,n}\right)^T = \nabla \left[(\mathbf{g}^{[\sigma]}(\mu))^{i_1...i_s}\right].$$

We compute the p-th entry in the above vector. On one hand, using Equation (30), we get:

$$(\nabla \mathbf{g}^{[\sigma]}(\mu))^{i_1...i_s,p} = \sum_{\substack{l=1 \\ i_l=p}}^{s} g^{[\sigma_{(l)}]}(\mu_{i_1}, ..., \mu_{i_s}, \mu_{i_l}).$$

On the other, using Equation (11), we evaluate the right-hand side of (32):

$$\left(\sum_{l=1}^{s}(T_l(\mu))^{\tau_l}\right)^{i_1...i_s,p} = \sum_{l=1}^{s}\left((T_l(\mu))^{\tau_l}\right)^{i_1...i_s,p}$$

$$= \sum_{l=1}^{s}(T_l(\mu))^{i_1...i_s}\delta_{i_l p}$$

$$= \sum_{\substack{l=1 \\ i_l=p}}^{s}(T_l(\mu))^{i_1...i_s}$$

$$= \sum_{\substack{l=1 \\ i_l=p}}^{s} g^{[\sigma_{(l)}]}(\mu_{i_1}, ..., \mu_{i_s}, \mu_{i_l}).$$
∎

We now continue the evaluation of $\nabla^s G(\text{Diag}\,\mu)[M]$. Using Theorem 2.5 in the last equality below, we find that

$$U\left(\sum_{\sigma \in \tilde{P}^s} \text{Diag}^{\sigma}\left(\nabla g^{[\sigma]}(\mu)[h]\right)\right)U^T = U\left(\sum_{\sigma \in \tilde{P}^s} \text{Diag}^{\sigma}\left(\left(\sum_{l=1}^{s}(T_l(\mu))^{\tau_l}\right)[h]\right)\right)U^T$$

17

$$= U\bigg(\sum_{\sigma\in\tilde{P}s}\mathrm{Diag}^{\sigma}\bigg(\sum_{l=1}^{s}\big(T_l(\mu)\big)^{\tau_l}[h]\bigg)\bigg)U^T$$

$$= U\bigg(\sum_{\sigma\in\tilde{P}s}\sum_{l=1}^{s}\mathrm{Diag}^{\sigma}\big(\big(T_l(\mu)\big)^{\tau_l}[h]\big)\bigg)U^T$$

$$= \sum_{\sigma\in\tilde{P}s}\sum_{l=1}^{s}U\bigg(\mathrm{Diag}^{\sigma}\big(\big(T_l(\mu)\big)^{\tau_l}[h]\big)\bigg)U^T$$

(33)
$$= \sum_{\sigma\in\tilde{P}s}\sum_{l=1}^{s}\Big(\mathrm{Diag}^{\sigma_{(l)}}\big(T_l(\mu)\big)_{\mathrm{in}}^{\tau_l}\Big)[M].$$

This already shows that $\nabla^{(s-1)}G(\mathrm{Diag}\,\mu)$ is differentiable. All that is left to do, now, is to show that $\nabla^s G(\mathrm{Diag}\,\mu)$ has the desired form and properties. The last step is formulated in the next lemma.

Lemma 6.5 *For every $l \in \mathbb{N}_s$ the following identity holds*

$$\mathbf{g}^{[\sigma_{(l)}]}(\mu) = \big(T_l(\mu)\big)_{\mathrm{in}}^{\tau_l} + \big(\mathbf{g}^{[\sigma]}(\mu)\big)_{\mathrm{out}}^{(l)}.$$

Proof. Fix a number $l \in \mathbb{N}_s$ and a multi index $(i_1, ..., i_s, i_{s+1})$. We consider two cases depending on whether or not $\mu_{i_{s+1}}$ equals μ_{i_l}.

Case I. Suppose $i_l \sim_\mu i_{s+1}$. Then, the entry of the left-hand side, corresponding to the multi index $(i_1, ..., i_s, i_{s+1})$ is

$$\big(\mathbf{g}^{[\sigma_{(l)}]}(\mu)\big)^{i_1...i_s i_{s+1}} = g^{[\sigma_{(l)}]}(\mu_{i_1}, ..., \mu_{i_s}, \mu_{i_{s+1}}) = \nabla_l g^{[\sigma]}(\mu_{i_1}, ..., \mu_{i_s}).$$

On the other hand, the right-hand side evaluates to

$$\begin{aligned}
\big(\big(T_l(\mu)\big)_{\mathrm{in}}^{\tau_l} + \big(\mathbf{g}^{[\sigma]}(\mu)\big)_{\mathrm{out}}^{(l)}\big)^{i_1...i_s i_{s+1}} &= \big(\big(T_l(\mu)\big)_{\mathrm{in}}^{\tau_l}\big)^{i_1...i_s i_{s+1}} + \big(\big(\mathbf{g}^{[\sigma]}(\mu)\big)_{\mathrm{out}}^{(l)}\big)^{i_1...i_s i_{s+1}} \\
&= \big(\big(T_l(\mu)\big)_{\mathrm{in}}^{\tau_l}\big)^{i_1...i_s i_{s+1}} + 0 \\
&= \big(T_l(\mu)\big)^{i_1...i_s} \\
&= g^{[\sigma_{(l)}]}(\mu_{i_1}, ..., \mu_{i_s}, \mu_{i_l}) \\
&= \nabla_l g^{[\sigma]}(\mu_{i_1}, ..., \mu_{i_s}),
\end{aligned}$$

where in the third equality we used Equation (10) and the fact that $T_l(\mu)$ is block-constant.

Case II. Suppose $i_l \not\sim_\mu i_{s+1}$. Then, the entry of the left-hand side, corresponding to the multi index $(i_1, ..., i_s, i_{s+1})$ is

$$\begin{aligned}
\big(\mathbf{g}^{[\sigma_{(l)}]}(\mu)\big)^{i_1...i_s i_{s+1}} &= g^{[\sigma_{(l)}]}(\mu_{i_1}, ..., \mu_{i_s}, \mu_{i_{s+1}}) \\
&= \frac{g^{[\sigma]}(\mu_{i_1}, ..., \mu_{i_l}, ..., \mu_{i_s}) - g^{[\sigma]}(\mu_{i_1}, ..., \mu_{i_{s+1}}, ..., \mu_{i_s})}{\mu_{i_l} - \mu_{i_{s+1}}},
\end{aligned}$$

18

where both μ_{i_l} and $\mu_{i_{s+1}}$ are in position l. On the other hand, the right-hand side evaluates to

$$
\begin{aligned}
\left(\left(T_l(\mu)\right)_{\text{in}}^{\tau_l} + \left(\mathbf{g}^{[\sigma]}(\mu)\right)_{\text{out}}^{(l)}\right)^{i_1...i_s i_{s+1}} &= \left(\left(T_l(\mu)\right)_{\text{in}}^{\tau_l}\right)^{i_1...i_s i_{s+1}} + \left(\left(\mathbf{g}^{[\sigma]}(\mu)\right)_{\text{out}}^{(l)}\right)^{i_1...i_s i_{s+1}} \\
&= 0 + \left(\left(\mathbf{g}^{[\sigma]}(\mu)\right)_{\text{out}}^{(l)}\right)^{i_1...i_s i_{s+1}} \\
&= \frac{\left(\mathbf{g}^{[\sigma]}(\mu)\right)^{i_1...i_{l-1} i_{s+1} i_{l+1}...i_s} - \left(\mathbf{g}^{[\sigma]}(\mu)\right)^{i_1...i_{l-1} i_l i_{l+1}...i_s}}{\mu_{i_{s+1}} - \mu_{i_l}} \\
&= \frac{g^{[\sigma]}\left(\mu_{i_1}, ..., \mu_{i_{s+1}}, ..., \mu_{i_s}\right) - g^{[\sigma]}\left(\mu_{i_1}, ..., \mu_{i_l}, ..., \mu_{i_s}\right)}{\mu_{i_{s+1}} - \mu_{i_l}}.
\end{aligned}
$$

In both cases, the two sides are equal and we are done. $\blacksquare$

Putting Equations (29) and (33) together, and using Lemma 6.5 concludes the inductive step and proves Theorem 6.1.

In the special case when matrix X has distinct eigenvalues, it seems that Theorem 5.1 and Theorem 6.1 give two different formulae for the higher-order derivatives of a separable spectral function. We now reconcile the differences. Suppose we have the formula for $\nabla^s G(X)$ given by Equation (26) and apply to it the inductive procedure described in Theorem 5.1 to obtain $\nabla^{(s+1)} G(X)$. The calculations in Subsection 6.3 showed that the gradient $\mathcal{A}_{\sigma_{(s+1)}} = \nabla \mathcal{A}_\sigma$ can be partitioned into s pieces (Lemma 6.4) and each piece can be added as an s-dimensional "diagonal plane" (Lemma 6.5) to a corresponding tensor $\mathcal{A}_{\sigma_{(l)}}$ for $l \in \mathbb{N}_s$. Doing that, we will arrive at the formula for $\nabla^{(s+1)} G(X)$ given by Theorem 6.1.

6.4 C^k separable spectral functions

Theorem 6.1 holds for every k-times differentiable functions g. If in addition g in k-times continuously differentiable, then Formula (26) can be significantly simplified. This is what we will describe in this section. In particular, we will show three properties of the functions $g^{[\sigma]}(x_1, ..., x_s)$, for every $2 \leq s \leq k+1$ and every $\sigma \in \tilde{P}_s$. First, we will give a compact determinant formula for computing $g^{[\sigma]}(x_1, ..., x_s)$ directly. Second, as a consequence of the determinant formula we will see that $g^{[\sigma]}(x_1, ..., x_s)$ is a symmetric function on its s arguments. Finally, third, denoting $\sigma_s = (12...s)$, all functions $g^{[\sigma]}(x_1, ..., x_s)$ can be obtained from $g^{[\sigma_s]}(x_1, ..., x_s)$ by a permutation of its arguments. (Thus, knowing one of the tensors in Formula (26), namely $\mathbf{g}^{[\sigma_s]}(\mu)$, we can obtain the rest by permuting its "rows" and "columns".)

Denote by $V(x_1, ..., x_s)$ the Vandermonde determinant

$$
V(x_1, ..., x_s) = \begin{vmatrix} x_1^{s-1} & x_2^{s-1} & \cdots & x_s^{s-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_1 & x_2 & \cdots & x_s \\ 1 & 1 & \cdots & 1 \end{vmatrix} = \prod_{j<i}(x_j - x_i).
$$

For any $y \in \mathbb{R}^s$, denote by $V\left(\begin{smallmatrix} y_1,\ldots,y_s \\ x_1,\ldots,x_s \end{smallmatrix}\right)$ the determinant

$$
V\left(\begin{smallmatrix} y_1,\ldots,y_s \\ x_1,\ldots,x_s \end{smallmatrix}\right) = \begin{vmatrix} y_1 & y_2 & \cdots & y_s \\ x_1^{s-2} & x_2^{s-2} & \cdots & x_s^{s-2} \\ \vdots & \vdots & \ddots & \vdots \\ x_1 & x_2 & \cdots & x_s \\ 1 & 1 & \cdots & 1 \end{vmatrix}.
$$

Lemma 6.6 *For any vector $(x_1, \ldots, x_s, x_{s+1})$ with distinct coordinates, any $y \in \mathbb{R}^{s+1}$, and $l \in \mathbb{N}_s$ the following identity holds*

$$
\frac{V\left(\begin{smallmatrix} y_1,\ldots,y_s \\ x_1,\ldots,x_s \end{smallmatrix}\right)}{V(x_1, \ldots, x_s)} - \frac{V\left(\begin{smallmatrix} y_1,\ldots,y_{l-1},y_{s+1},y_{l+1},\ldots,y_s \\ x_1,\ldots,x_{l-1},x_{s+1},x_{l+1},\ldots,x_s \end{smallmatrix}\right)}{V(x_1, \ldots, x_{l-1}, x_{s+1}, x_{l+1}, \ldots, x_s)} = (x_l - x_{s+1}) \frac{V\left(\begin{smallmatrix} y_1,\ldots,y_l,y_{s+1},y_{l+1},\ldots,y_s \\ x_1,\ldots,x_l,x_{s+1},x_{l+1},\ldots,x_s \end{smallmatrix}\right)}{V(x_1, \ldots, x_l, x_{s+1}, x_{l+1}, \ldots, x_s)}.
$$

Proof. We consider both sides of the above identity as a multivariate polynomial in the variables $y_1, \ldots, y_s, y_{s+1}$ and show that the coefficients in front of y_k on both sides are equal for all $k \in \mathbb{N}_{s+1}$. Notice first that

$$
V(x_1, \ldots, x_{l-1}, x_{s+1}, x_{l+1}, \ldots, x_s) = (-1)^{s-l} V(x_1, \ldots, x_{l-1}, x_{l+1}, \ldots, x_s, x_{s+1}),
$$
$$
V\left(\begin{smallmatrix} y_1,\ldots,y_{l-1},y_{s+1},y_{l+1},\ldots,y_s \\ x_1,\ldots,x_{l-1},x_{s+1},x_{l+1},\ldots,x_s \end{smallmatrix}\right) = (-1)^{s-l} V\left(\begin{smallmatrix} y_1,\ldots,y_{l-1},y_{l+1},\ldots,y_s,y_{s+1} \\ x_1,\ldots,x_{l-1},x_{l+1},\ldots,x_s,x_{s+1} \end{smallmatrix}\right).
$$

We consider four cases according to the partition $\mathbb{N}_{s+1} = \{1, \ldots, l-1\} \cup \{l\} \cup \{l+1, \ldots, s\} \cup \{s+1\}$. (In all product formulae below, it will be assumed that the index $j < i$. This conditions is omitted for typographical reasons. Also a hat on top of a multiple in a product denotes that the multiple is missing.) First, let $k \in \{1, \ldots, l-1\}$. The coefficient in front of y_k in the expression on the left-hand side is equal to

$$
(-1)^{k+1} \frac{\prod_{i,j \in \mathbb{N}_{s+1} \backslash \{k,s+1\}} (x_j - x_i)}{\prod_{i,j \in \mathbb{N}_{s+1} \backslash \{s+1\}} (x_j - x_i)} - (-1)^{k+1} \frac{\prod_{i,j \in \mathbb{N}_{s+1} \backslash \{k,l\}} (x_j - x_i)}{\prod_{i,j \in \mathbb{N}_{s+1} \backslash \{l\}} (x_j - x_i)}
$$

$$
= \frac{(-1)^{k+1}}{(x_1 - x_k) \cdots (x_{k-1} - x_k)(x_k - x_{k+1}) \cdots (x_k - x_s)}
$$

$$
- \frac{(-1)^{k+1}}{(x_1 - x_k) \cdots (x_{k-1} - x_k)(x_k - x_{k+1}) \cdots \widehat{(x_k - x_l)} \cdots (x_k - x_{s+1})}
$$

$$
= \frac{(-1)^{k+1}}{(x_1 - x_k) \cdots (x_{k-1} - x_k)(x_k - x_{k+1}) \cdots \widehat{(x_k - x_l)} \cdots (x_k - x_s)} \left(\frac{1}{x_k - x_l} - \frac{1}{x_k - x_{s+1}} \right)
$$

$$
= \frac{(-1)^{k+1}(x_l - x_{s+1})}{(x_1 - x_k) \cdots (x_{k-1} - x_k)(x_k - x_{k+1}) \cdots (x_k - x_{s+1})}
$$

$$
= (-1)^{k+1}(x_l - x_{s+1}) \frac{\prod_{i,j \in \mathbb{N}_{s+1} \backslash \{k\}} (x_j - x_i)}{\prod_{i,j \in \mathbb{N}_{s+1}} (x_j - x_i)},
$$

which is the coefficient in front of y_k on the right-hand side of the identity.

Suppose now, $k = l$. Then, the coefficient of y_l in the left-hand side of the identity is

$$(-1)^{l+1}\frac{\prod_{i,j\in\mathbb{N}_{s+1}\setminus\{l,s+1\}}(x_j - x_i)}{\prod_{i,j\in\mathbb{N}_{s+1}\setminus\{s+1\}}(x_j - x_i)} - 0 = \frac{(-1)^{l+1}}{(x_1 - x_l)\cdots(x_{l-1} - x_l)(x_l - x_{l+1})\cdots(x_l - x_s)}$$

$$= \frac{(-1)^{l+1}(x_l - x_{s+1})}{(x_1 - x_l)\cdots(x_{l-1} - x_l)(x_l - x_{l+1})\cdots(x_l - x_{s+1})}$$

$$= (-1)^{l+1}(x_l - x_{s+1})\frac{\prod_{i,j\in\mathbb{N}_{s+1}\setminus\{l\}}(x_j - x_i)}{\prod_{i,j\in\mathbb{N}_{s+1}}(x_j - x_i)}.$$

When $k \in \{l+1, ..., s\}$, the coefficient in front of y_k in the left-hand side of the identity is:

$$(-1)^{k+1}\frac{\prod_{i,j\in\mathbb{N}_{s+1}\setminus\{k,s+1\}}(x_j - x_i)}{\prod_{i,j\in\mathbb{N}_{s+1}\setminus\{s+1\}}(x_j - x_i)} - (-1)^{k+2}\frac{\prod_{i,j\in\mathbb{N}_{s+1}\setminus\{k,l\}}(x_j - x_i)}{\prod_{i,j\in\mathbb{N}_{s+1}\setminus\{l\}}(x_j - x_i)}$$

$$= \frac{(-1)^{k+1}}{(x_1 - x_k)\cdots(x_{k-1} - x_k)(x_k - x_{k+1})\cdots(x_k - x_s)}$$

$$- \frac{(-1)^{k+2}}{(x_1 - x_k)\cdots(\widehat{x_l - x_k})\cdots(x_{k-1} - x_k)(x_k - x_{k+1})\cdots(x_k - x_{s+1})}$$

$$= \frac{(-1)^{k+1}}{(x_1 - x_k)\cdots(\widehat{x_l - x_k})\cdots(x_{k-1} - x_k)(x_k - x_{k+1})\cdots(x_k - x_s)}\left(\frac{1}{x_l - x_k} + \frac{1}{x_k - x_{s+1}}\right)$$

$$= \frac{(-1)^{k+1}(x_l - x_{s+1})}{(x_1 - x_k)\cdots(x_{k-1} - x_k)(x_k - x_{k+1})\cdots(x_k - x_{s+1})}$$

$$= (-1)^{k+1}(x_l - x_{s+1})\frac{\prod_{i,j\in\mathbb{N}_{s+1}\setminus\{k\}}(x_j - x_i)}{\prod_{i,j\in\mathbb{N}_{s+1}}(x_j - x_i)},$$

which is the coefficient in front of y_k on the right-hand side of the identity.

Finally, when $k = s+1$ the coefficient of y_{s+1} in the left-hand side of the identity is

$$0 - (-1)^{l+1}(-1)^{s-l}\frac{\prod_{i,j\in\mathbb{N}_{s+1}\setminus\{l,s+1\}}(x_j - x_i)}{\prod_{i,j\in\mathbb{N}_{s+1}\setminus\{l\}}(x_j - x_i)} = \frac{(-1)^{s+2}}{(x_1 - x_{s+1})\cdots(\widehat{x_l - x_{s+1}})\cdots(x_s - x_{s+1})}$$

$$= \frac{(-1)^{s+2}(x_l - x_{s+1})}{(x_1 - x_{s+1})\cdots(x_s - x_{s+1})}$$

$$= (-1)^{s+2}(x_l - x_{s+1})\frac{\prod_{i,j\in\mathbb{N}_{s+1}\setminus\{s+1\}}(x_j - x_i)}{\prod_{i,j\in\mathbb{N}_{s+1}}(x_j - x_i)},$$

which is again the coefficient of y_{s+1} on the right. ∎

21

Theorem 6.7 *Suppose $g \in C^k(I)$. Then, for every permutation $\sigma \in \tilde{P}^s$, $2 \le s \le k+1$, and every vector $(x_1, ..., x_s)$ with distinct coordinates, we have the formula*

(34)
$$g^{[\sigma]}(x_1, ..., x_s) = \frac{V\left(\begin{smallmatrix} g(x_1),...,g(x_s) \\ x_1,...,x_s \end{smallmatrix}\right)}{V(x_1, ..., x_s)}.$$

In particular, $g^{[\sigma]}(x_1, ..., x_s)$ is symmetric everywhere in its domain.

Proof. The proof is by induction on s. When $s = 2$ and $x_1 \ne x_2$, then by the definition we have the representation

$$g^{[(12)]}(x_1, x_2) = \frac{\begin{vmatrix} g(x_1) & g(x_2) \\ 1 & 1 \end{vmatrix}}{\begin{vmatrix} x_1 & x_2 \\ 1 & 1 \end{vmatrix}} = \frac{V(g(x_1), g(x_2); x_1, x_2)}{V(x_1, x_2)}.$$

Suppose Representation (34) holds for s, $2 \le s < k+1$. Fix a permutation $\sigma \in \tilde{P}^s$ and an $l \in \mathbb{N}_s$, then $\sigma_{(l)} \in \tilde{P}^{s+1}$. Let $y = (g(x_1), ..., g(x_s), g(x_{s+1}))$. Using Definition (24) for any point $(x_1, ..., x_s, x_{s+1})$ with distinct coordinates together with Lemma 6.6 and the induction hypothesis, we get

$$g^{[\sigma_{(l)}]}(x_1, ..., x_s, x_{s+1}) = \frac{g^{[\sigma]}(x_1, ..., x_s) - g^{[\sigma]}(x_1, ..., x_{l-1}, x_{s+1}, x_{l+1}, ..., x_s)}{x_l - x_{s+1}}$$

$$= \frac{1}{(x_l - x_{s+1})} \left(\frac{V\left(\begin{smallmatrix} y_1,...,y_s \\ x_1,...,x_s \end{smallmatrix}\right)}{V(x_1, ..., x_s)} - \frac{V\left(\begin{smallmatrix} y_1,...,y_{l-1},y_{s+1},y_{l+1},...,y_s \\ x_1,...,x_{l-1},x_{s+1},x_{l+1},...,x_s \end{smallmatrix}\right)}{V(x_1, ..., x_{l-1}, x_{s+1}, x_{l+1}, ..., x_s)} \right)$$

$$= \frac{V\left(\begin{smallmatrix} y_1,...,y_l,y_{s+1},y_{l+1},...,y_s \\ x_1,...,x_l,x_{s+1},x_{l+1},...,x_s \end{smallmatrix}\right)}{V(x_1, ..., x_l, x_{s+1}, x_{l+1}, ..., x_s)}$$

$$= \frac{V\left(\begin{smallmatrix} y_1,...,y_{s+1} \\ x_1,...,x_{s+1} \end{smallmatrix}\right)}{V(x_1, ..., x_{s+1})}.$$

Since $\tilde{P}^{s+1} = \{\sigma_{(l)} \mid \sigma \in \tilde{P}^s, l \in \mathbb{N}_s\}$ the induction step is completed. Finally, using the continuity of $g^{[\sigma]}(x_1, ..., x_s)$ shows that it is a symmetric function everywhere on its domain. ∎

Now, we simplify Theorem 6.1 significantly. Define the $(k+1)$-tensor valued map

(35)
$$\mathbf{g} : \mathbb{R}^n \to T^{k+1,n}, \text{ where}$$

$$\left(\mathbf{g}(\mu)\right)^{i_1...i_{k+1}} := \frac{V\left(\begin{smallmatrix} g(\mu_{i_1}),...,g(\mu_{i_{k+1}}) \\ \mu_{i_1},...,\mu_{i_{k+1}} \end{smallmatrix}\right)}{V(\mu_{i_1}, ..., \mu_{i_{k+1}})}.$$

Technically, this definition is good only at μ's with distinct coordinates, but Lemma 6.7 shows that it can be extended continuously everywhere. Clearly, if $(i_1, ..., i_{k+1}) \sim_\mu (j_1, ..., j_{k+1})$, then $\left(\mathbf{g}(\mu)\right)^{i_1...i_{k+1}} = \left(\mathbf{g}(\mu)\right)^{j_1...j_{k+1}}$, which shows that $\mathbf{g}(\mu)$ is a μ-block-constant tensor for every μ. Moreover, $\mathbf{g}(\mu)$ is a symmetric tensor, and the map $\mathbf{g} : \mathbb{R}^n \to T^{k+1,n}$ is continuous.

Theorem 6.8 *Let g be a C^k function defined on an interval I. Let X be a symmetric matrix with eigenvalues in the interval I, and let V be an orthogonal matrix such that $X = V(\operatorname{Diag}\lambda(X))V^T$. Then, the matrix valued function G defined by (21) and (22) is C^k at X. We have the formula*

(36)
$$\nabla^k G(X) = V\left(\sum_{\sigma \in \tilde{P}^{k+1}} \operatorname{Diag}^{\sigma} \mathbf{g}(\lambda(X))\right)V^T,$$

where the $(k+1)$-tensor valued maps $\mathbf{g}(\cdot)$ is defined by Equation (35).

The next corollary is a generalization of Formula (V.22) from [2]. It is a specialization of the last theorem to the case when $k = 2$. Since G is a symmetric matrix valued function, the second derivative $\nabla^2 G(\operatorname{Diag}\mu)[H_1, H_2]$ can be viewed as a symmetric matrix. For every $i = 1, ..., n$, define the projection onto the i-th coordinate axis

$$P_i : \mathbb{R}^n \to \mathbb{R}^n$$
$$P_i(x) = x_i e^i.$$

Corollary 6.9 *For $g \in C^2(I)$ and any $n \times n$ symmetric matrices H_1, H_2, H_3 we have*

$$\langle \nabla^2 G(X)[H_1, H_2], H_3 \rangle = 2 \sum_{p_1,p_2,p_3=1}^{n,n,n} \mathbf{g}(\lambda(X))^{p_1p_2p_3} \tilde{H}_1^{p_1p_3} \tilde{H}_2^{p_2p_1} \tilde{H}_3^{p_3p_2},$$

$$\nabla^2 G(X)[H_1, H_2] = 2 \sum_{p_1,p_2,p_3=1}^{n,n,n} \mathbf{g}(\lambda(X))^{p_1p_2p_3} P_{p_1} \tilde{H}_1 P_{p_2} \tilde{H}_2 P_{p_3},$$

where $X = V(\operatorname{Diag}\lambda(X))V^T$, and $\tilde{H}_i = V^T H_i V$, $i = 1, 2, 3$.

Proof. Suppose first that $X = \operatorname{Diag}\mu$ for some $\mu \in \mathbb{R}^n_{\downarrow}$.

$$\langle \nabla^2 G(\operatorname{Diag}\mu)[H_1, H_2], H_3 \rangle = \nabla^2 G(\operatorname{Diag}\mu)[H_1, H_2, H_3]$$
$$= \left(\sum_{\sigma \in \tilde{P}^3} \operatorname{Diag}^{\sigma} \mathbf{g}(\mu)\right)[H_1, H_2, H_3]$$
$$= \sum_{\sigma \in \tilde{P}^3} \langle \mathbf{g}(\mu), H_1 \circ_{\sigma} H_2 \circ_{\sigma} H_3 \rangle$$
$$= \langle \mathbf{g}(\mu), H_1 \circ_{(123)} H_2 \circ_{(123)} H_3 \rangle + \langle \mathbf{g}(\mu), H_1 \circ_{(132)} H_2 \circ_{(132)} H_3 \rangle$$
$$= \sum_{p_1,p_2,p_3=1}^{n,n,n} \mathbf{g}(\mu)^{p_1p_2p_3} H_1^{p_1p_3} H_2^{p_2p_1} H_3^{p_3p_2} + \sum_{q_1,q_2,q_3=1}^{n,n,n} \mathbf{g}(\mu)^{q_1q_2q_3} H_1^{q_1q_2} H_2^{q_2q_3} H_3^{q_3q_1}.$$

After re-parametrization of the second sum ($p_1 = q_2$, $p_2 = q_3$, $p_3 = q_1$), and using the fact that $\mathbf{g}(\mu)$ is a symmetric tensor, we continue

$$= \sum_{p_1,p_2,p_3=1}^{n,n,n} (\mathbf{g}(\mu)^{p_1p_2p_3} + \mathbf{g}(\mu)^{p_3p_1p_2}) H_1^{p_1p_3} H_2^{p_2p_1} H_3^{p_3p_2} = 2 \sum_{p_1,p_2,p_3=1}^{n,n,n} \mathbf{g}(\mu)^{p_1p_2p_3} H_1^{p_1p_3} H_2^{p_2p_1} H_3^{p_3p_2}.$$

23

To show the second representation of $\nabla^2 G(\text{Diag}\,\mu)[H_1, H_2]$ and the general case, when X is not an ordered diagonal matrix, is routine. $\blacksquare$

7 The Hessian of spectral functions, revisited

In this last section, we illustrate one more time the machinery developed so far. We recalculate the Hessian of a general spectral functions at an arbitrary matrix.

One of the strengths of the new approach is that one doesn't need to have a preconceived notion about the form of the these derivatives. (Recall that in [6] the formula for the Hessian of the spectral function was first stated and, then it was proven that is indeed the correct one. The hind sight for that formula came from [7].) Here, we simply differentiate applying the rules developed so far to arrive at the correct formula. The approach also clearly shows where the different pieces of the Hessian come from. This should make the calculation routine and more clear.

7.1 Two matrix valued maps

Let $\mu \in \mathbb{R}^n \to T(\mu) \in T^{1,n}$, be a μ-symmetric, differentiable, 1-tensor-valued map. (In the next section, $T(\mu) = \nabla f(\mu)$, where f is asymmetric C^2 function.) We define two matrix valued maps $D_0 T$ and $D_1 T$ that play an important role in the description of the Hessian of spectral functions. First

$$D_0 T(\mu) = \nabla T(\mu),$$

or in other words

$$D_0 T(\mu)^{i_1 i_2} = \frac{\partial}{\partial \mu_{i_2}} (T(\mu)^{i_1}).$$

Next, define the matrix $D_1 T(\mu)$ as follows

$$D_1 T(\mu)^{i_1 i_2} = \begin{cases} 0, & \text{if } i_1 = i_2, \\ (\nabla T)^{i_1 i_1}(\mu) - (\nabla T)^{i_1 i_2}(\mu), & \text{if } i_1 \sim i_2, \\ \dfrac{T^{i_2}(\mu) - T^{i_1}(\mu)}{\mu_{i_2} - \mu_{i_1}}, & \text{if } i_1 \not\sim i_2, \end{cases}$$

where the equivalence relation is with respect to the vector μ. Several of the properties of $D_1 T$ are easily seen from the following integral representations.

Lemma 7.1 *If $T(\mu) \in T^{1,n}$ is continuously differentiable, and μ-symmetric map, then for every i_1, $i_2 \in \{1, ..., n\}$ we have the representation*

$$D_1 T(\mu)^{i_1 i_2} = \int_0^1 (\nabla T)^{i_1 i_1}(\cdots, \mu_{i_1} + t(\mu_{i_2} - \mu_{i_1}), \cdots, \mu_{i_2} + t(\mu_{i_1} - \mu_{i_2}), \cdots) -$$

$$(\nabla T)^{i_1 i_2}(\cdots, \mu_{i_1} + t(\mu_{i_2} - \mu_{i_1}), \cdots, \mu_{i_2} + t(\mu_{i_1} - \mu_{i_2}), \cdots)dt,$$

where the first displayed argument is in position i_1 and the second displayed argument is in position i_2. The missing arguments are the corresponding, unchanged, entries of μ.

Proof. The first case, when $i_1 = i_2$ is immediate. In the second, $i_l \sim i_2$ implies that $\mu_{i_1} = \mu_{i_2}$ and the integrand doesn't depend on t. In the third case, $i_1 \not\sim i_2$, we can compute the integral using the Fundamental Theorem of Calculus:

$$D_l T(\mu)^{i_1 i_2} = \frac{1}{\mu_{i_2} - \mu_{i_1}} \int_0^1 \frac{\partial}{\partial t} T^{i_1}\Big(..., \mu_{i_1} + t(\mu_{i_2} - \mu_{i_1}), ..., \mu_{i_2} + t(\mu_{i_1} - \mu_{i_2}), ...\Big)\, dt$$

$$= \frac{T^{i_1}(..., \mu_{i_2}, ..., \mu_{i_1}, ...) - T^{i_1}(..., \mu_{i_1}, ..., \mu_{i_2}, ...)}{\mu_{i_2} - \mu_{i_1}}$$

$$= \frac{T^{i_2}(..., \mu_{i_1}, ..., \mu_{i_2}, ...) - T^{i_1}(..., \mu_{i_1}, ..., \mu_{i_2}, ...)}{\mu_2 - \mu_{i_1}},$$

where the last equality follows from the fact that $T(\mu)$ is μ-symmetric. ∎

Lemma 7.2 *If $T(\mu)$ is differentiable, then both $D_0 T(\mu)$ and $D_1 T(\mu)$ are μ-symmetric maps.*

Proof. The fact that $D_0 T(\mu)$ is μ-symmetric is Lemma 2.1. This implies that if $i_1 \sim j_1$, then $(\nabla T)^{i_1 i_1}(\mu) = (\nabla T)^{j_1 j_1}(\mu)$. Also, if $i_1 \sim j_1$ and $i_2 \sim j_2$ with $i_1 \neq i_2$ and $j_1 \neq j_2$, then $(\nabla T)^{i_1 i_2}(\mu) = (\nabla T)^{j_1 j_2}(\mu)$. The fact that T is μ-symmetric implies that if $i_1 \sim j_1$, then $T^{i_1}(\mu) = T^{j_1}(\mu)$. Now it is easy to see that $D_1 T(\mu)$ is μ-symmetric. ∎

We conclude this section with a summary of the properties of $D_0 T(\mu)$ and $D_1 T(\mu)$

- For every $l = 0, 1$, $D_l T(\mu)$ is a matrix valued, μ-symmetric map.
- For every $l = 0, 1$, $D_l T(\mu)$ is as smooth as $\nabla T(\mu)$. In other words, if $\nabla T(\mu)$ is continuous, or several times (continuously) differentiable, then so is $D_l T(\mu)$.
- In addition, if $T = \nabla f(\mu)$ where $f : \mathbb{R}^n \to \mathbb{R}$ is a symmetric C^2 function, then for every $l = 0, 1$, $D_l T(\mu)$ is a symmetric matrix for every μ.

7.2 $f \circ \lambda$ is twice (continuously) differentiable if, and only if, f is

Suppose that f is a symmetric function, twice differentiable at $\mu \in \mathbb{R}^n_\downarrow$. Let E be an arbitrary symmetric matrix. Using Formula (19) together with Formula (16) we compute:

$$\lim_{m \to \infty} \frac{\nabla(f \circ \lambda)(\operatorname{Diag}\mu + M_m) - \nabla(f \circ \lambda)(\operatorname{Diag}\mu)}{\|M_m\|}$$

$$= \lim_{m \to \infty} \frac{U_m\big(\operatorname{Diag}{}^{(1)}\nabla f(\lambda(\operatorname{Diag}\mu + M_m)))\big)U_m^T - \operatorname{Diag}{}^{(1)}\nabla f(\mu)}{\|M_m\|}$$

$$= \lim_{m \to \infty} \frac{U_m \big(\mathrm{Diag}^{\,(1)} \nabla f(\mu + h_m + o(\|M_m\|))\big) U_m^T - \mathrm{Diag}^{\,(1)} \nabla f(\mu)}{\|M_m\|}$$

$$= \lim_{m \to \infty} \frac{U_m \big(\mathrm{Diag}^{\,(1)} (\nabla f(\mu) + \nabla^2 f(\mu)[h_m] + o(\|M_m\|))\big) U_m^T - \mathrm{Diag}^{\,(1)} \nabla f(\mu)}{\|M_m\|}$$

$$= \lim_{m \to \infty} \frac{U_m \big(\mathrm{Diag}^{\,(1)} (\nabla f(\mu))\big) U_m^T - \mathrm{Diag}^{\,(1)} \nabla f(\mu)}{\|M_m\|} + U \big(\nabla^2 f(\mu)[h]\big) U^T.$$

For convenience let $T = \nabla f(\mu)$, a block-constant 1-tensor. Using Corollary 2.4 we see that

$$\lim_{m \to \infty} \frac{U_m \big(\mathrm{Diag}^{\,(1)} T\big) U_m^T - \mathrm{Diag}^{\,(1)} T}{\|M_m\|} = \big(\mathrm{Diag}^{\,(12)} T_{\mathrm{out}}^{(1)}\big)[M].$$

Denote $\mathcal{A}_1 = D_0 T$, where the operator D_0 is defined in Section 7.1. Notice that there is a block-constant vector b such that $\mathcal{A}_1 - \mathrm{Diag}\, b$ is a block-constant 2-tensor. Using this notation and Corollary 2.5 we continue:

$$U \big(\nabla^2 f(\mu)[h]\big) U^T = U \big((\mathcal{A}_1 - \mathrm{Diag}\, b + \mathrm{Diag}\, b)[h]\big) U^T$$

$$= U \big((\mathcal{A}_1 - \mathrm{Diag}\, b)[h]\big) U^T + U \big((\mathrm{Diag}\, b)[h]\big) U^T$$

$$= \big(\mathrm{Diag}^{\,(1)(2)} (\mathcal{A}_1 - \mathrm{Diag}\, b)\big)[M] + \big(\mathrm{Diag}^{\,(12)} b_{\mathrm{in}}^{(1)}\big)[M].$$

This, shows that $f \circ \lambda$ is twice differentiable.

In order to prove that $f \circ \lambda$ is twice continuously differentiable we need to reorganize the pieces. Let $\mathcal{A}_2 = D_1 T$, where the operator D_1 is defined in Section 7.1. Notice that the sum $\mathcal{A}_1 + \mathcal{A}_2$ is block-constant 2-tensor. This means that vector b is (can be chosen) such that $\mathcal{A}_2 + \mathrm{Diag}\, b$ is block-constant, and

$$\mathcal{A}_2 + \mathrm{Diag}\, b = T_{\mathrm{out}}^{(1)} + b_{\mathrm{in}}^{(1)}.$$

Putting everything together we obtain:

$$\nabla^2 (f \circ \lambda)(\mathrm{Diag}\, \mu) = \mathrm{Diag}^{\,(12)} T_{\mathrm{out}}^{(1)} + \mathrm{Diag}^{\,(1)(2)} (\mathcal{A}_1 - \mathrm{Diag}\, b) + \mathrm{Diag}^{\,(12)} b_{\mathrm{in}}^{(1)}$$

$$= \mathrm{Diag}^{\,(1)(2)} (\mathcal{A}_1 - \mathrm{Diag}\, b) + \mathrm{Diag}^{\,(12)} (\mathcal{A}_2 + \mathrm{Diag}\, b)$$

$$= \mathrm{Diag}^{\,(1)(2)} \mathcal{A}_1 + \mathrm{Diag}^{\,(12)} \mathcal{A}_2.$$

In the last equality we used the fact that $\mathrm{Diag}^{\,(1)(2)}(\mathrm{Diag}\, b) = \mathrm{Diag}^{\,(12)}(\mathrm{Diag}\, b)$, which is very easy to verify. The discussion in [8, Section 6] shows that

$$(37) \qquad \nabla^2 (f \circ \lambda)(X) = V \big(\mathrm{Diag}^{\,(1)(2)} \mathcal{A}_1(\lambda(X)) + \mathrm{Diag}^{\,(12)} \mathcal{A}_2(\lambda(X))\big) V^T,$$

where $X = V(\mathrm{Diag}\, \lambda(X)) V^T$.

Moreover, we showed in Section 7.1 that if f is C^2, then both $\mathcal{A}_1$ and $\mathcal{A}_2$ are continuous. By Proposition 6.2 in [8] it follows that $\nabla^2 (f \circ \lambda)$ is continuous. That is, f is C^2 if, and only if, $f \circ \lambda$ is.

References

[1] V.I. ARNOLD. On matrices depending on parameters. *Russian Mathematical Surveys*, 26:29–43, 1971.

[2] R. BHATIA. *Matrix Analysis.* Springer-Verlag, NY, 1997.

[3] J.-B. HIRIART-URRUTY and D. YE. Sensitivity analysis of all eigenvalues of a symmetric matrix. *Numerische Mathematik*, 70:45–72, 1992.

[4] A.S. LEWIS. Derivatives of spectral functions. *Mathematics of Operations Research*, 21:576–588, 1996.

[5] A.S. LEWIS. Nonsmooth analysis of eigenvalues. *Mathematical Programming*, 84:1–24, 1999.

[6] A.S. LEWIS and H.S. SENDOV. Twice differentiable spectral functions. *SIAM Journal of Matrix Analysis and Applications*, 23(2):368–386, 2001.

[7] A.S. LEWIS and H.S. SENDOV. Quadratic expansions of spectral functions. *Linear Algebra and Its Applications*, 340:97–121, 2002.

[8] H.S. SENDOV. Generalized Hadamard product and the derivatives of spectral functions. *Research Report, Department of Mathematics and Statistics, University of Guelph, Ontario*, 2004. Mathematical Series 2004-308.

[9] H.S. SENDOV. Interactions between the generalized Hadamard product and the eigenvalues of symmetric matrices. *Research Report, Department of Mathematics and Statistics, University of Guelph, Ontario*, 2004. Mathematical Series 2004-309.

[10] D. SUN and J. SUN. Semismooth matrix-valued functions. *Mathematics of Operations Research*, 27(1):150–169, 2002.

[11] N.-K. TSING, M. K. H. FAN, and E. I. VERRIEST. On analyticity of functions involving eigenvalues. *Linear Algebra and its Applications*, 207:159–180, 1994.